

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平11-45245

(43)公開日 平成11年(1999)2月16日

(51)Int.Cl.⁶

G 0 6 F 17/28

識別記号

F I

G 0 6 F 15/38

P

T

審査請求 未請求 請求項の数13 FD (全 10 頁)

(21)出願番号 特願平9-217129

(22)出願日 平成9年(1997)7月27日

(71)出願人 390024350

株式会社ジャストシステム
徳島県徳島市沖浜東3-46

(72)発明者 野村 直之

徳島県徳島市沖浜東3丁目46番地 株式会
社ジャストシステム内

(74)代理人 弁理士 川井 隆 (外1名)

(54)【発明の名称】 外国語文読解支援システム、外国語文読解支援プログラムが記憶された記憶媒体、及び外国語文
読解支援方法

(57)【要約】

【課題】 より自然な翻訳結果を得ることが可能な外国
語文読解支援システムを提供する。

【解決手段】 まず最初に翻訳対象文書（英語等）か
ら、自然言語処理により複合名詞句等の所定の非訳出語
句を抽出する。そして非訳出語句の部分を逐語訳するこ
となく、カタカナ表記に変換して対象文書全体を翻訳す
る。このようにすることで、例えば「Action Manager」
を変に逐語訳せずに、「アクション マネージャ」とカ
タカナ表記することで、その内容を知っている文章全体
としての読解が容易になる。また、内容を知らないユー
ザであっても、意味不明の逐語訳よりも、特殊な語句で
あることを前提に他の部分を読むことができるため混乱
することなく読み進むことができる。

原文	日本語
Action Manager is now varing. So	アクション マネージャ ーは、今や多様化してい る。 従って、.....
...	...

【特許請求の範囲】

【請求項1】 翻訳対象となる対象文書を取得する対象文書取得手段と、この対象文書取得手段で取得された対象文書から非訳出語句を抽出する非訳出語句抽出手段と、翻訳辞書と、前記翻訳辞書を使用して前記対象文書を翻訳する翻訳手段とを備え、前記翻訳手段は、前記非訳出語句抽出手段で抽出された非訳出語句の訳語として前記対象文書中の原文表記を訳語として翻訳することを特徴とする外国語文読解支援システム。

【請求項2】 翻訳対象となる対象文書を取得する対象文書取得手段と、前記対象文書取得手段で取得された対象文書から非訳出語句を抽出する非訳出語句抽出手段と、前記非訳出語句抽出手段で抽出された非訳出語句を翻訳対象語による表記に変換する表記変換手段と、翻訳辞書と、前記翻訳辞書を使用して前記対象文書を翻訳する翻訳手段とを備え、前記翻訳手段は、前記非訳出語句抽出手段で抽出された非訳出語句の訳語として、前記表記変換手段により変換された表記、または、前記表記変換手段により変換された表記と前記対象文書中の原文表記の両表記を使用して、前記対象文書を翻訳することを特徴とする外国語文読解支援システム。

【請求項3】 前記非訳出語句抽出手段は、前記翻訳辞書中には独立して存在しない名詞句を非訳出語句として抽出する、ことを特徴とする請求項1又は請求項2に記載の外国語文読解支援システム。

【請求項4】 前記非訳出語句抽出手段は、複合名詞句を非訳出語句として抽出する、ことを特徴とする請求項1又は請求項2に記載の外国語文読解支援システム。

【請求項5】 前記非訳出語句の重要度を判定する重要度判定手段を備え、

前記翻訳手段は、この重要度判定手段により判定された重要度に応じて前記非訳出語句を他の翻訳部分と区別して表示することを特徴とする請求項1から請求項4のうちのいずれか1の請求項に記載の外国語文読解支援システム。

【請求項6】 前記対象文書取得手段により取得された対象文書の段落、文、及び非訳出語句のうちの少なくとも1つの単位に対する類似文書または関連記述を検索する検索手段と、

前記検索手段で検索された類似文書または関連記述を、該当単位に関連付ける関連付け手段と、を具備することを特徴とする請求項1から請求項5のうちのいずれか1の請求項に記載の外国語文読解支援システム。

【請求項7】 翻訳対象となる対象文書を取得する対象

文書取得機能と、この対象文書取得機能で取得された対象文書から非訳出語句を抽出する非訳出語句抽出機能と、

翻訳辞書と、前記翻訳辞書を使用して前記対象文書を翻訳する翻訳機能とを備え、前記翻訳機能は、前記非訳出語句抽出機能で抽出された非訳出語句の訳語として前記対象文書中の原文表記を訳語として翻訳することをコンピュータに実現させるためのコンピュータ読み取り可能な外国語文読解支援プログラムが記憶された記憶媒体。

【請求項8】 翻訳対象となる対象文書を取得する対象文書取得機能と、前記対象文書取得機能で取得された対象文書から非訳出語句を抽出する非訳出語句抽出機能と、前記非訳出語句抽出機能で抽出された非訳出語句を翻訳対象語による表記に変換する表記変換機能と、翻訳辞書と、

前記翻訳辞書を使用して前記対象文書を翻訳する翻訳機能とを備え、前記翻訳機能は、前記非訳出語句抽出機能で抽出された非訳出語句の訳語として、前記表記変換機能により変換された表記、または、前記表記変換機能により変換された表記と前記対象文書中の原文表記の両表記を使用して、前記対象文書を翻訳することをコンピュータに実現させるためのコンピュータ読み取り可能な外国語文読解支援プログラムが記憶された記憶媒体。

【請求項9】 前記非訳出語句抽出機能は、前記翻訳辞書中には独立して存在しない名詞句を非訳出語句として抽出する、ことを特徴とする請求項7又は請求項8に記載の外国語文読解支援プログラムが記憶された記憶媒体。

【請求項10】 前記非訳出語句抽出機能は、複合名詞句を非訳出語句として抽出する、ことを特徴とする請求項7又は請求項8に記載の外国語文読解支援プログラムが記憶された記憶媒体。

【請求項11】 前記非訳出語句の重要度を判定する重要度判定機能を備え、

前記翻訳機能は、この重要度判定機能により判定された重要度に応じて前記非訳出語句を他の翻訳部分と区別して表示することを特徴とする請求項7から請求項10のうちのいずれか1の請求項に記載の外国語文読解支援プログラムが記憶された記憶媒体。

【請求項12】 前記対象文書取得機能により取得された対象文書の段落、文、及び非訳出語句のうちの少なくとも1つの単位に対する類似文書または関連記述を検索する検索機能と、

前記検索機能で検索された類似文書または関連記述を、該当単位に関連付ける関連付け機能と、を具備することを特徴とする請求項7から請求項11のうちのいずれか

1の請求項に記載の外国語文読解支援プログラムが記憶された記憶媒体。

【請求項13】 翻訳辞書を使用して前記対象文書を翻訳する際に、

翻訳対象となる対象文書から非訳出語句を抽出し、抽出した非訳出語句の訳語として、翻訳対象語による表記に変換した非訳出語句の表記と、前記対象文書中の原文表とのうちの少なくとも一方の表記を使用して翻訳することを特徴とする外国語文読解支援方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は、外国語文読解支援システム、外国語文読解支援プログラムが記憶された記憶媒体、及び外国語文読解支援方法に係り、詳細には、取得した文書の翻訳に関する。

【0002】

【従来の技術】 外国語で記載された文書を日本語の文書に翻訳するシステムのように、所定の言語で作成された文書（外国語文）をそれと異なる言語の文書に翻訳する翻訳システムが実用化されている。翻訳システムは、インターネットやパソコン通信等の各種通信により自由に且つ大量に情報を入手できるようになった現在では、広く使用されるようになってきている。この翻訳システムでは、翻訳用の辞書を使用して原文（翻訳対象文書）を一定のアルゴリズムに従って所定の言語に翻訳している。翻訳は比較的短時間で行われるため、通信を行っている場合に外国語で作成されたホームページを理解するうえで有効である。

【0003】

【発明が解決しようとする課題】 しかし、従来の翻訳システムでは、辞書中に存在しない語句が原文中にあると、その語句を逐語訳しているため、文全体としてよくわからない翻訳になったり、時には読み解不能になる場合もあった。これは、本来一体として意味をなすべき複合名詞句等の語句を正確に認識できていないにもかかわらず、これらの語句を分離して逐語訳をしてしまうためである。例えば、図13に示すように、「Action Manager is now varing. So…」といった原文に対して、従来の翻訳システムでは、翻訳辞書に存在しない「Action Manager」の部分を逐語訳により「演技課長」と翻訳してしまう。その結果、「演技課長は、今や多様化している。従って…」という翻訳文では何のことかよくわからなかった。

【0004】 そこで本発明は、前記従来の課題を解決するために成されたもので、より自然な翻訳結果を得ることが可能な外国語文読解支援システムを提供することを第1の目的とする。また本発明は、より自然な翻訳結果を得ることが可能な外国語文読解支援プログラムが記憶された記憶媒体を提供することを第2の目的とする。また本発明は、より自然な翻訳結果を得ることが可能な外

国語文読解支援方法を提供することを第3の目的とする。

【0005】

【課題を解決するための手段】 請求項1に記載した外国語文読解支援システムでは、図5に示すように、翻訳対象となる対象文書を取得する対象文書取得手段501と、この対象文書取得手段501で取得された対象文書から非訳出語句を抽出する非訳出語句抽出手段502と、翻訳辞書503と、前記翻訳辞書503を使用して

10 前記対象文書を翻訳する翻訳手段504とを備え、前記翻訳手段504は、前記非訳出語句抽出手段502で抽出された非訳出語句の訳語として前記対象文書中の原文表記を訳語として翻訳することにより、前記第1の目的を達成する。請求項2に記載した外国語文読解支援システムでは、図6に示すように、翻訳対象となる対象文書を取得する対象文書取得手段501と、前記対象文書取得手段501で取得された対象文書から非訳出語句を抽出する非訳出語句抽出手段502と、前記非訳出語句抽出手段502で抽出された非訳出語句を翻訳対象語による表記に変換する表記変換手段505と、翻訳辞書503と、前記翻訳辞書503を使用して前記対象文書を翻訳する翻訳手段506とを備え、前記翻訳手段506は、前記非訳出語句抽出手段502で抽出された非訳出語句の訳語として、前記表記変換手段505により変換された表記、または、前記表記変換手段505により変換された表記と前記対象文書中の原文表記の両表記を使用して、前記対象文書を翻訳することにより、前記第1の目的を達成する。請求項3に記載した発明では、図6、図7に示すように、請求項1又は請求項2に記載した外国語文読解支援システムにおいて、前記非訳出語句抽出手段505は、前記翻訳辞書503中には独立して存在しない名詞句を非訳出語句として抽出する。請求項4に記載した発明では、図6、図7に示すように、請求項1又は請求項2に記載した外国語文読解支援システムにおいて、前記非訳出語句抽出手段502は、複合名詞句を非訳出語句として抽出する。請求項5に記載した発明では、図7に1例を示すように、請求項1から請求項4のうちのいずれか1の請求項に記載の外国語文読解支援システムにおいて、前記非訳出語句の重要度を判定する重要度判定手段507を備え、前記翻訳手段504

20 は、この重要度判定手段507により判定された重要度に応じて前記非訳出語句を他の翻訳部分と区別して表示する。請求項6に記載した発明では、図7に1例を示すように、請求項1から請求項4のうちのいずれか1の請求項に記載の外国語文読解支援システムにおいて、前記対象文書取得手段501により取得された対象文書の段落、文、及び非訳出語句のうちの少なくとも1つの単位に対する類似文書または関連記述を検索する検索手段508と、前記検索手段508で検索された類似文書または関連記述を、該当単位に関連付ける関連付け手段50

30 3と、前記翻訳辞書503を使用して前記対象文書を翻訳する翻訳手段506とを備え、前記翻訳手段506は、前記非訳出語句抽出手段502で抽出された非訳出語句の訳語として、前記表記変換手段505により変換された表記、または、前記表記変換手段505により変換された表記と前記対象文書中の原文表記の両表記を使用して、前記対象文書を翻訳することにより、前記第1の目的を達成する。請求項3に記載した発明では、図6、図7に示すように、請求項1又は請求項2に記載した外国語文読解支援システムにおいて、前記非訳出語句抽出手段505は、前記翻訳辞書503中には独立して存在しない名詞句を非訳出語句として抽出する。請求項4に記載した発明では、図6、図7に示すように、請求項1又は請求項2に記載した外国語文読解支援システムにおいて、前記非訳出語句抽出手段502は、複合名詞句を非訳出語句として抽出する。請求項5に記載した発明では、図7に1例を示すように、請求項1から請求項4のうちのいずれか1の請求項に記載の外国語文読解支援システムにおいて、前記非訳出語句の重要度を判定する重要度判定手段507を備え、前記翻訳手段504

40 は、この重要度判定手段507により判定された重要度に応じて前記非訳出語句を他の翻訳部分と区別して表示する。請求項6に記載した発明では、図7に1例を示すように、請求項1から請求項4のうちのいずれか1の請求項に記載の外国語文読解支援システムにおいて、前記対象文書取得手段501により取得された対象文書の段落、文、及び非訳出語句のうちの少なくとも1つの単位に対する類似文書または関連記述を検索する検索手段508と、前記検索手段508で検索された類似文書または関連記述を、該当単位に関連付ける関連付け手段50

9と、を具備させる。請求項7に記載した発明では、図9に示すように、翻訳対象となる対象文書を取得する対象文書取得機能601と、この対象文書取得機能601で取得された対象文書から非訳出語句を抽出する非訳出語句抽出機能602と、翻訳辞書603と、前記翻訳辞書603を使用して前記対象文書を翻訳する翻訳機能604とを備え、前記翻訳機能604は、前記非訳出語句抽出機能602で抽出された非訳出語句の訳語として前記対象文書中の原文表記を訳語として翻訳することをコンピュータに実現させるためのコンピュータ読み取り可能な外国語文読解支援プログラムが記憶された記憶媒体として、前記第2の目的を達成する。請求項8に記載した発明では、図10に示すように、翻訳対象となる対象文書を取得する対象文書取得機能601と、前記対象文書取得機能601で取得された対象文書から非訳出語句を抽出する非訳出語句抽出機能602と、前記非訳出語句抽出機能602で抽出された非訳出語句を翻訳対象語による表記に変換する表記変換機能605と、翻訳辞書603と、前記翻訳辞書603を使用して前記対象文書を翻訳する翻訳機能604とを備え、前記翻訳機能604は、前記非訳出語句抽出機能602で抽出された非訳出語句の訳語として、前記表記変換機能605により変換された表記、または、前記表記変換機能605により変換された表記と前記対象文書中の原文表記の両表記を使用して、前記対象文書を翻訳することをコンピュータに実現させるためのコンピュータ読み取り可能な外国語文読解支援プログラムが記憶された記憶媒体として、前記第2の目的を達成する。請求項9に記載した発明では、図9、図10に示すように、請求項7または請求項8に記載した外国語文読解支援プログラムが記憶された記憶媒体において、前記非訳出語句抽出機能602は、前記翻訳辞書中には独立して存在しない名詞句を非訳出語句として抽出する。請求項10に記載した発明では、図9、図10に示すように、請求項7または請求項8に記載した外国語文読解支援プログラムが記憶された記憶媒体において、前記非訳出語句抽出機能602は、複合名詞句を非訳出語句として抽出する。請求項11に記載した発明では、図11に1例を示すように、請求項7から請求項10のうちのいずれか1の請求項に記載の外国語文読解支援プログラムが記憶された記憶媒体において、前記非訳出語句の重要度を判定する重要度判定機能607を備え、前記翻訳機能604は、この重要度判定機能607により判定された重要度に応じて前記非訳出語句を他の翻訳部分と区別して表示する。請求項12に記載した発明では、図12に1例を示すように、請求項7から請求項11のうちのいずれか1の請求項に記載の外国語文読解支援プログラムが記憶された記憶媒体において、前記対象文書取得機能601により取得された対象文書の段落、文、及び非訳出語句のうちの少なくとも1つの単位に対する類似文書または関連記述を検

索する検索機能608と、前記検索機能608で検索された類似文書または関連記述を、該当単位に関連付ける関連付け機能609と、を具備させる。請求項13に記載した外国語文読解支援方法では、翻訳辞書を使用して前記対象文書を翻訳する際に、翻訳対象となる対象文書から非訳出語句を抽出701し、抽出した非訳出語句の訳語として、翻訳対象語による表記に変換した非訳出語句の表記と、前記対象文書中の原文表とのうちの少なくとも一方の表記を使用して翻訳702する。

【0006】

【発明の実施の形態】以下、本発明の外国語文読解支援システム、外国語文読解支援プログラムが記憶された記憶媒体、及び外国語文読解支援方法における好適な実施の形態について、図1から図9を参照して説明する。

(1) 実施形態の概要

本実施形態による外国語文読解支援は、まず最初に翻訳対象文書（英語等）から、自然言語処理により複合名詞句等の所定の非訳出語句を抽出する。そして非訳出語句の部分を逐語訳すことなく、カタカナ表記に変換して

20 対象文書全体を翻訳する。このようにすることで、例えば「Action Manager」を変に逐語訳せずに、「アクション マネージャ」とカタカナ表記することで、その内容を知っている文章全体としての読解が容易になる。また、内容を知らないユーザであっても、意味不明の逐語訳よりも、特殊な語句であることを前提に他の部分を読むことができるため混乱することなく読み進むことができる。

【0007】(2) 実施の形態の詳細

本実施形態の外国語文読解支援システムは、パソコン30 コンピュータやワードプロセッサ等を含むコンピュータシステムで構成するだけでなく、LAN（ローカル・エリア・ネットワーク）のサーバ、コンピュータ（パソコン）通信のホスト、インターネット上に接続されたコンピュータシステム等によって構成することも可能である。また、ネットワーク上の各機器に機能分散させ、ネットワーク全体で外国語文読解支援システムを構成することも可能である。

【0008】図1は、外国語文読解支援システムの構成を表したブロック図である。外国語文読解支援システム40 は、図1に示すようにシステム全体を制御するための制御部11を備えている。この制御部11には、データバス等のバスライン21を介して、入力装置としてのキーボード12やマウス13、表示装置14、印刷装置15、記憶装置16、記憶媒体駆動装置17、通信制御装置18、入出力I/F19、及び文字認識装置20が接続されている。制御部11は、CPU111、ROM112、RAM113を備えている。ROM112は、CPU111が各種制御や演算を行うための各種プログラムやデータが予め格納されたリードオンリーメモリである。

【0009】RAM113は、CPU111にワーキングメモリとして使用されるランダムアクセスメモリである。このRAM113には、本実施形態による外国語文読解支援処理を行うためのエリアとして、翻訳対象となる対象文書が格納される対象文書格納エリア1131と、対象文書から翻訳しない語句として抽出した非訳出語句とそのカタカナ表記を格納する非訳出語句格納エリア1132と、非訳出語句にカタカナ表記を使用して翻訳した翻訳文書を格納する翻訳文書格納エリア1133その他の各種エリアが確保されるようになっている。

【0010】キーボード12は、自システム内で対象文書や他言語を入力する場合の対象文書取得手段や他言語入力手段の一部を構成し、かな文字を入力するためのかなキーやテンキー、各種機能を実行するための機能キー、カーソルキー、等の各種キーが配置されている。マウス13は、ポインティングデバイスであり、表示装置14に表示されたキーやアイコン等を左クリックすることで対応する機能の指定を行う入力装置である。表示装置14は、例えばCRTや液晶ディスプレイ等が使用される。この表示装置には、キーボード12やマウス13による入力結果が表示されたり、外国語文読解支援処理における、対象文書や翻訳文書等が表示されたりするようになっている。印刷装置15は、表示装置14に表示された文書や、記憶装置16の文書格納部164に格納された文書等や、ユーザに対する請求書の印刷を行うためのものである。この印刷装置としては、レーザプリンタ、ドットプリンタ、インクジェットプリンタ、ページプリンタ、感熱式プリンタ、熱転写式プリンタ、等の各種印刷装置が使用される。

【0011】記憶装置16は、読み書き可能な記憶媒体と、その記憶媒体に対してプログラムやデータ等の各種情報を読み書きするための駆動装置で構成されている。この記憶装置16に使用される記憶媒体としては、主としてハードディスクが使用されるが、後述の記憶媒体駆動装置17で使用される各種記憶媒体のうちの読み書き可能な記憶媒体を使用するようにしてもよい。記憶装置16は、仮名漢字変換辞書161、プログラム格納部162、データ格納部163、文書データベース164、アクセス履歴データベース165、翻訳に使用される翻訳辞書165、図示しないその他の格納部（例えば、この記憶装置16内に格納されているプログラムやデータ等をバックアップするための格納部）等を有している。プログラム格納部162には、本実施形態における外国語文読解支援処理プログラム、非訳出語句抽出プログラム等の各種プログラムの他、仮名漢字変換辞書161を使用して入力された仮名文字列を漢字混り文に変換する仮名漢字変換プログラム等の各種プログラムが格納されている。データ格納部163には、ユーザに関するデータ等の、システムが必要とする各種データが格納されている。文書データベース164には、主として日本語の

文書が多数格納されている。必要に応じてこの日本語文書のなかから所定の文書等と類似している類似文書が検索される。この文書データベース164に格納される各文書の形式は特に限定されるものではなく、テキスト形式の文書、HTML(Hyper Text Markup Language)形式の文書、JIS形式の文書等の各種形式の文書の格納が可能である。

【0012】記憶媒体駆動装置17は、CPU111が外部の記憶媒体からコンピュータプログラムや文書を含むデータ等を読み込むための駆動装置である。記憶媒体に記憶されているコンピュータプログラム等には、本実施形態の外国語文読解支援システムにより実行される外国語文読解支援処理等の各種処理プログラム、および、そこで使用される辞書、データ等も含まれる。ここで、記憶媒体とは、コンピュータプログラムやデータ等が記憶される記憶媒体をいい、具体的には、フロッピーディスク、ハードディスク、磁気テープ等の磁気記憶媒体、メモリチップやICカード等の半導体記憶媒体、CD-ROMやMO、PD(相変化書換型光ディスク)等の光学的に情報が読み取られる記憶媒体、紙カードや紙テープ等の用紙（および、用紙に相当する機能を持った媒体）を用いた記憶媒体、その他各種方法でコンピュータプログラム等が記憶される記憶媒体が含まれる。本実施形態の外国語文読解支援システムにおいて使用される記憶媒体としては、主として、CD-ROMやフロッピーディスクが使用される。記憶媒体駆動装置17は、これらの各種記憶媒体からコンピュータプログラムを読み込む他に、フロッピーディスクのような書き込み可能な記憶媒体に対してRAM113や記憶装置16に格納されているデータ等を書き込むことが可能である。

【0013】本実施形態の外国語文読解支援システムでは、制御部11のCPU111が、記憶媒体駆動装置17にセットされた外部の記憶媒体からコンピュータプログラムを読み込んで、記憶装置16の各部に格納（インストール）する。そして、本実施形態による外国語文読解支援処理等の各種処理を実行する場合、記憶装置16から該当プログラムをRAM113に読み込み、実行するようになっている。但し、記憶装置16からではなく、記憶媒体駆動装置17により外部の記憶媒体から直接RAM113にプログラムを読み込んで実行することも可能である。また、外国語文読解支援システムによっては、本実施形態の外国語文読解支援処理プログラム等を予めROM112に記憶させておき、これをCPU111が実行するようにもよい。さらに、本実施形態の外国語文読解支援処理プログラム等の各種プログラムやデータを、通信制御装置18を介して他の記憶媒体からダウンロードし、実行するようにもよい。

【0014】通信制御装置18は、外国語文読解支援システムと他のパーソナルコンピュータやワードプロセッサ等の各種電子機器との間をネットワーク接続するため

の制御装置である。通信制御装置18は、これら各種電子機器が有している対象文書と同一の言語の文書、入力された他言語の文書、および同一言語や他言語の文書のデータベースを検索対象としてアクセスすることが可能になっている。対象となる文書には、テキスト形式やHTML形式等の各種形式の文書の他、ピットマップデータ等の各種データも含まれる。入出力I/F19は、音声や音楽等の出力をを行うスピーカ等の各種機器を接続するためのインターフェースである。文字認識装置20は、用紙等に記載された文字をテキスト形式やHTML等の各種形式で認識する装置であり、イメージスキャナや文字認識プログラム等で構成されている。

【0015】本実施形態では、キーボード12の入力操作により作成され文書(RAM113の所定格納エリアに格納)の他、記憶媒体駆動装置17により記憶媒体から読み出した外部で作成された文書、予め文書データベースに格納されている文書、通信制御装置18からダウンロードした文書、及び文字認識装置20で文字認識した文書、等の各種文書を翻訳の対象となる対象文書として取得する(文書取得手段)ことが可能である。

【0016】以上のように構成された本実施形態の外国語文読解支援システムによる外国語文読解支援処理の動作について、図2から図4を使用して説明する。本実施形態では、翻訳を希望する文書(原文)が英語で記載されており、これを日本語に翻訳する場合を例に説明する。

【0017】図2は外国語文読解支援処理のメイン動作を表したフローチャートである。CPU111は、まず外国語文読解支援(翻訳)を希望する対象文書(原文)を、取得し(ステップ11)。

【0018】そしてキーボード12やマウス13等の入力装置で指定された原文の言語(例えば英語)と翻訳言語(例えば日本語)、及び、表示形式を取得し、RAM113の所定格納エリアに格納する(ステップ12)。ここで、ユーザにより原文言語の指定がされない場合、CPU111は、取得した対象文書を解析して言語を特定する。解析は、各言語固有の表記文字の存在をチェックすることで行う。また、翻訳言語のユーザ選択は原則として必須項目であるため、選択がされない場合には、画面表示により選択を促す。ただし、外国語文読解支援システムが1種類の翻訳(例えば、英語から日本語のみ)だけが可能である場合には、原文言語と翻訳言語の取得と格納は不要である。CPU111は、原文言語、翻訳言語、表示形式を取得する場合、翻訳可能な言語の組み合わせと、表示形式を画面に表示してユーザに選択してもらうようとする。

【0019】次にCPU111は、RAM113の対象文書格納エリア1132に格納した対象文書から、原文言語に対応した非訳出語句抽出プログラムを起動し、非訳出語句を資源言語処理(NLP)抽出し、抽出した

非訳出語句(原文の表記)を非訳出語句格納エリア1132に格納する(ステップ13)。ここで、非訳出語句として抽出する語句の例としては、複合名詞句が抽出され、その例として、名詞+代名詞、名詞+固有名詞、固有名詞+冠詞+普通名詞(例: Billy the kid)、形容詞+普通名詞+固有名詞(例: Mad scientist Doctor Emmett Brown)等がある。更に例示すれば、原文言語が英語の場合、「An incremental utterance task-oriented dialogue production system」、「An Internet based wide area computer assisted foreign language learning prototype demonstration system」等の複合名詞句が非訳出語句として抽出される。

【0020】次にCPU111は、ステップ12で指定された表示形式が、カタカナ表示を含む表示形式か否かを確認する(ステップ14)。カタカナ表示を含む表示形式である場合には(; Y)、ステップ13で抽出した非訳出語句をカタカナ表記に変換し(表記変換手段)、非訳出語句に対応させて変換したカタカナ表記を非訳出語句格納エリア1132に格納する(ステップ15)。

20 ここで、非訳出語句のカタカナ表記への変換は、原文が英語の場合にはローマ字読みにより、または、各単語の発音記号により変換する。

【0021】そして、CPU111は、翻訳辞書165と、非訳出語句格納エリア1132に格納された非訳出語句及び/又はカタカナ表記(選択された表示形式によって選択)とを使用して対象文書を翻訳し翻訳結果を翻訳文書格納エリア1133に格納する(ステップ16)。翻訳が終了するとCPU111は、RAM113に格納した翻訳文書を指定された表示形式で表示して

30 (ステップ17)、処理を終了する。

【0022】図3は、表示形式として「カタカナ表記」が選択された場合の翻訳結果の表示画面を表したものである。この図3では、原文(対象文書)と日本語翻訳結果とを横に対比して表示しているが、いずれか一方のみを表示したり、上下に対比して表示したりすることも可能である。この図に示したように、非訳出語句「Action Manager」を変に「演技課長」といった意味不明の語に逐語訳せず、「アクション マネージャ」と一括してカタカナで表記しているので、当該言語を知っているユーザは直ちに理解することができる。また、当該言語を知らないユーザであっても、意味不明の逐語訳で文章全体がわからなくなるということなく、特殊な語句であることを前提に他の部分を読むことができるため混乱することなく読み進むことができる。

40 【0023】図4は、他の表示形式が選択された場合の表示画面を表したものであり、(A)は「原文表記」の表示画面、(B)は「カタカナ表記(原文表記)」の表示画面をあらわしたものである。この図に示すように、原文表記をすることで、その原文の語学力があるユーザにとっては原文表記から直ちに内容を理解することができ

き、そうでないユーザとては、原文表記から辞書等によって調べなければいけない単語を明確にすることができます。なお、図示しないが、表示形式として「原文表記（カタカナ表記）」が選択された場合には、非訳出語句「Action Manager」にたいして、「Action Manager（アクション マネージャ）」と表示する。

【0024】以上、本実施形態による対象文書に対してカタカナ表記／原文表記混りの翻訳を行う処理について説明したが、更に、以下の処理を行うようにしてもよい。すなわち、ステップ1-3で抽出した非訳出語句の重要度を対象文書を使用して判定し、その重要度に応じて、翻訳結果の翻訳文中の原文表記やカタカナ表記部分を他の部分特別して表示するようにする。原文表記やカタカナ表記の区別表示としては、色、濃度、文字サイズ、文字の太さ等の文字飾り情報等を用いて強調したり、逆強調したりするようにしてもよい。強調するのは重要度が高い場合で、赤等の目立つ色、高濃度、拡大文字（重要度に応じて大きさを変える）、太文字、のいずれか1以上により強調する。逆強調するのは重要度が低い場合で、水色や灰色等の目立たない色、低濃度、縮小文字、細文字、の1以上により目立たないように逆強調する。

【0025】非訳出語句の重要度の判定方法としては、各種提案されている公知の重要語判定方法を使用する。例えば、PU111は、形態素解析を行うことで対象文書から自立語を抽出して候補語とする。そして抽出した候補語の対象文書中の出現頻度、評価関数から、各候補語の重要度を決定し、一定の閾値以上の重要度を有する候補語を重要語とする。ここで、評価関数としては、例えば、所定の語句が予め指定されている場合にはその語句に対する重み付け等が使用される。そして、非訳出語句中に出現する重要語候補をチェックし、重要度の合*

$$\text{類似度 } S_{nn+1} = \cos(q) \\ = (b_n \cdot b_{n+1}) / (|b_n| \times |b_{n+1}|)$$

【0030】この類似度Sの値は $-1 \leq S \leq 1$ までの値をとり、1に近いほど2つの文書ベクトルが互いに平行に近く、2つの文書同士は似ていると考えることができる。従って、一定の閾値以上の類似度を有する文書を類似する文書と判断し、検索結果とする。

【0031】以上、本実施形態の構成および外国語文読解支援の処理について説明したが、本発明では、これらの各形態に限定されるものではなく、請求項に記載された発明の範囲内で種々の変形をすることが可能である。例えば、実施形態では、対象文書の言語（原文言語）として英語、翻訳言語として日本語を例に説明したが、本発明ではこれらの組み合わせに限定されるものではなく、あらゆる言語間の組み合わせとすることが可能である。その場合、対象となる文書が作成された言語用の形態素解析アルゴリズム等を使用するといった、本発明の構成には影響のない部分を変更するだけでよい。

* 計値をその非訳出語句の重要度とする。

【0026】また、非訳出語句の重要度の判定方法として、「全文情報からの意味的情報の抽出と加工」（情報処理学会第38回全国大会予稿集、第222頁；1989年）で提案されているように、文書中の重要語を字種や動詞等の情報から抽出して判定するようにしてもよい。

【0027】また、本実施形態による対象文書に対してカタカナ表記／原文表記混りの翻訳を行う処理に加えて、更に、以下の処理を行うようにしてもよい。すなわち、ステップ1-1で取得した対象文書の、段落、文、非訳出語句のうちの少なくとも1つ毎に、による類似文書または関連記述を検索し、ハイパーリンクを付ける。そして、ユーザーが当該箇所にスクロールしたりマウスカーソルをかざしたりした際に、バルーン（吹き出し）状に関連文書名一覧や文書内容の一部（先頭や要約）を表示し、理解を支援する。この際、母国語の関連文書を優先表示するようにしてもよい。

【0028】ここで、類似文書の検索は、公知の各種検索方法を使用することができ、例えば、文書の特徴を表す文書ベクトル（文書中に含まれるキーワードの重要度を要素とするベクトル）を使用し、2文書に対する文書ベクトル間の角度に依存するコサインにより以下のようにして両文書の類似度を求める。すなわち、両文書ベクトル b_n と b_{n+1} の角度を q とし、両文書ベクトルの内積を $b_n \cdot b_{n+1}$ とし、両文書ベクトルの大きさをそれぞれ $|b_n|$ 、 $|b_{n+1}|$ とした場合、両文書ベクトルの類似度 S_{nn+1} は次の式により求まる。

【0029】

【数1】

$$S_{nn+1} = (b_n \cdot b_{n+1}) / (|b_n| \times |b_{n+1}|)$$

【0032】以上の実施形態において説明した、各装置、各部、各動作、各処理等に対しては、それらを含む上位概念としての各手段（～手段）により、実施形態を構成することが可能である。例えば、「2文書に対する文書ベクトル間の角度に依存するコサインにより以下のようにして両文書の類似度を求める」との記載に対して「類似度算出手段」を構成するようにしてもよい。同様に、その他各種動作に対して「～（動作）手段」等の上位概念で実施形態を構成するようにしてもよい。

【0033】

【発明の効果】本発明によれば、翻訳辞書を使用して前記対象文書を翻訳する際に、翻訳対象となる対象文書から非訳出語句を抽出し、抽出した非訳出語句の訳語として、翻訳対象語による表記に変換した非訳出語句の表記と、前記対象文書中の原文表とのうちの少なくとも一方の表記を使用して翻訳するようにしたので、より自然な

翻訳結果が提供され、外国語文読解が支援される。

【図面の簡単な説明】

【図1】本発明の1実施形態における外国語文読解支援システムの構成を表したブロック図である。

【図2】同上、実施形態における外国語文読解支援処理のメイン動作を表したフローチャートである。

【図3】同上、実施形態における表示形式として「カタカナ表記」が選択された場合の翻訳結果の表示画面を表した説明図である。

【図4】同上、実施形態における他の表示形式が選択された場合の表示画面を表したもので、(A)は「原文表記」の表示画面、(B)は「カタカナ表記(原文表記)」の表示画面をあらわした説明図である。

【図5】請求項1に記載した発明のクレーム対応図である。

【図6】請求項2に記載した発明のクレーム対応図である。

【図7】請求項5に記載した発明のクレーム対応図の1例である。

【図8】請求項6に記載した発明のクレーム対応図の1例である。

【図9】請求項7に記載した発明のクレーム対応図である。

【図10】請求項8に記載した発明のクレーム対応図である。

【図11】請求項11に記載した発明のクレーム対応図の1例である。

* 【図12】請求項12に記載した発明のクレーム対応図の1例である。

【図13】請求項13に記載した発明のクレーム対応図である。

【図14】従来の翻訳結果を表す説明図である。

【符号の説明】

1 1 制御部

1 1 2 ROM

1 1 3 RAM

10 1 1 3 1 対象文書格納エリア

1 1 3 2 非訳出語句格納エリア

1 1 3 3 翻訳文書格納エリア

1 2 キーボード

1 3 マウス

1 4 表示装置

1 5 印刷装置

1 6 記憶装置

1 6 1 仮名漢字変換辞書

1 6 2 プログラム格納部

20 1 6 3 データ格納部

1 6 4 文書データベース

1 6 5 翻訳辞書

1 7 記憶媒体駆動装置

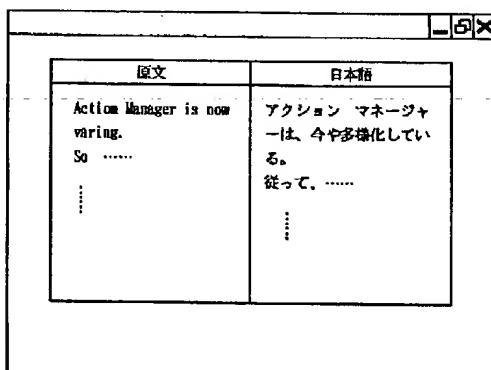
1 8 通信制御装置

1 9 入出力 I/F

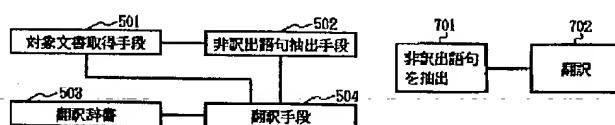
2 0 文字認識装置

*

【図3】



【図5】



【図13】

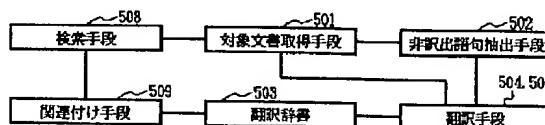
【図7】



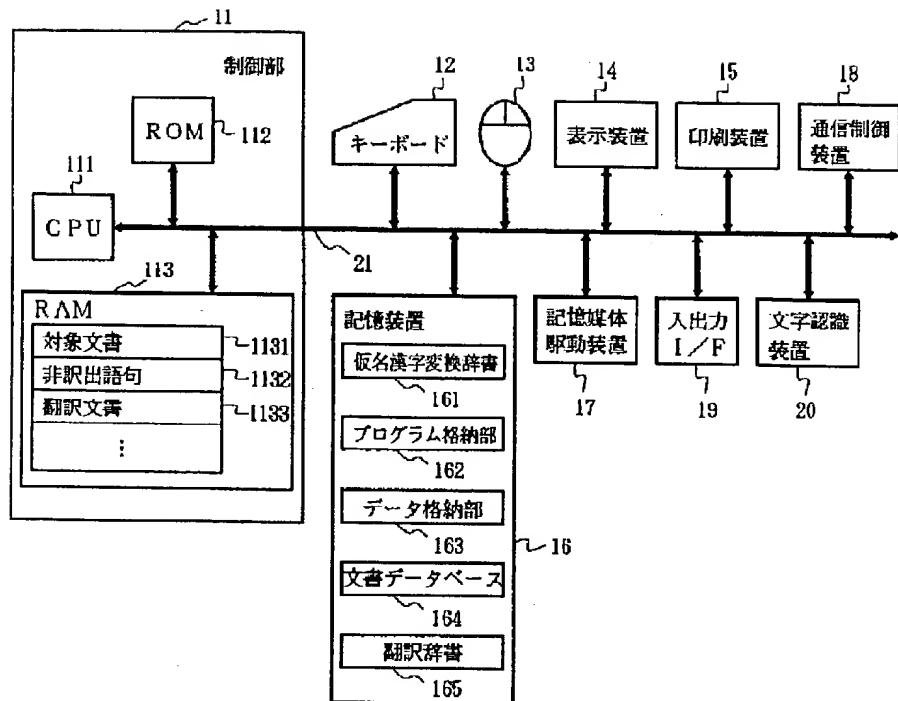
【図6】



【図8】

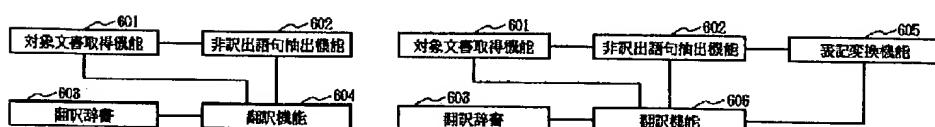


【図1】



【図9】

【図10】



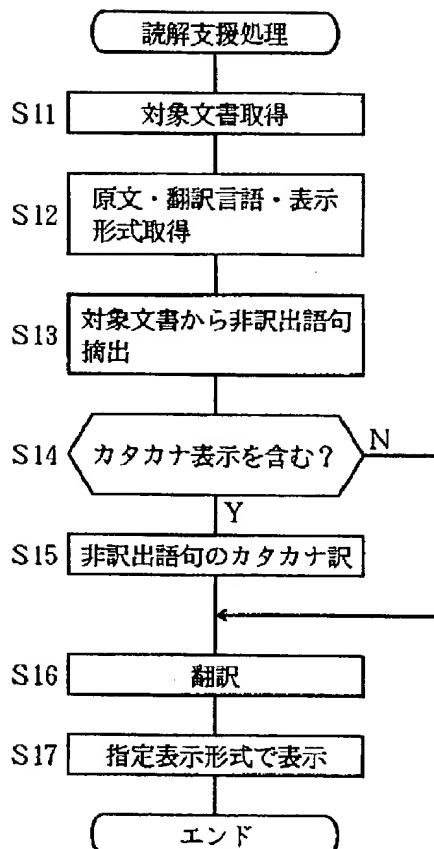
【図11】



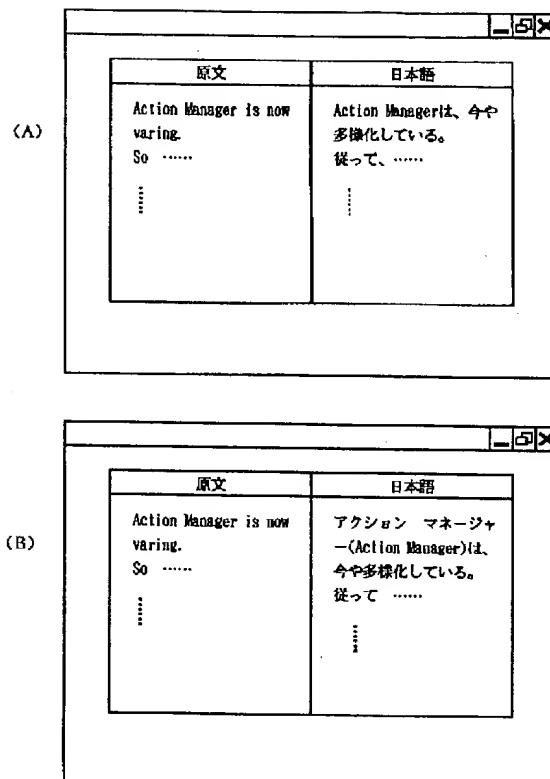
【図12】



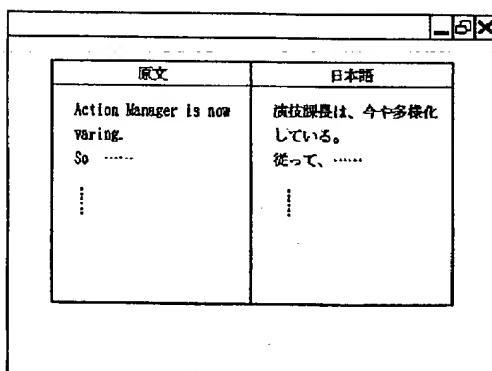
【図2】



【図4】



【図14】



PATENT ABSTRACTS OF JAPAN

(11)Publication number : 11-045245
(43)Date of publication of application : 16.02.1999

(51)Int.CI.

G06F 17/28

(21)Application number : 09-217129
(22)Date of filing : 27.07.1997

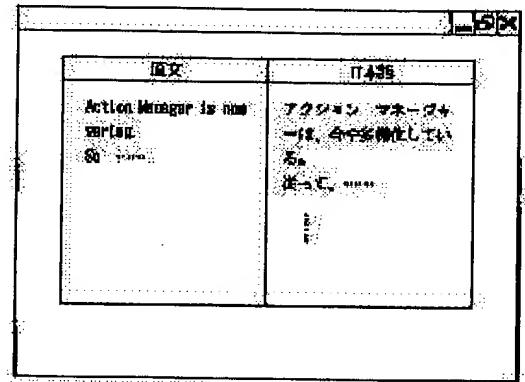
(71)Applicant : JUST SYST CORP
(72)Inventor : NOMURA NAOYUKI

(54) FOREIGN LANGUAGE SENTENCE INTERPRETATION SUPPORT SYSTEM, STORING MEDIUM FOR STORING FOREIGN LANGUAGE SENTENCE INTERPRETATION SUPPORT PROGRAM AND METHOD FOR FOREIGN LANGUAGE SENTENCE INTERPRETATION SUPPORT

(57)Abstract:

PROBLEM TO BE SOLVED: To provide a system for foreign language sentence interpretation support system that can obtain more natural translation results.

SOLUTION: At the outset, specific non-translated words and phrases such as compound noun phrase or the like are extracted from a translation object document (in English and so on) by a natural language processing. Then, the whole object document is translated by converting part of non-translated words and phrases to katakana Japanese syllabary transcription without translating the part of non-translated words and phrases word by word. Thus, reading as the whole text of which the contents are known becomes easy by transcribing katakana, for example 'AKUSHON MANEJA' instead of strangely translating 'Action Manager' word by word. Also, since even a user who does not know the contents can read other parts on the assumption that they are written in special words and phrases rather than an obscure verbatim translation, he can go on reading without being confused.



EGAL STATUS

Date of request for examination]

Date of sending the examiner's decision of rejection]

Kind of final disposal of application other than the examiner's decision of rejection or application
[converted registration]

Date of final disposal for application]

Patent number]

Date of registration]

Number of appeal against examiner's decision of ejection]

Date of requesting appeal against examiner's decision of rejection]

Date of extinction of right]

[JP,11-045245,A]

* NOTICES *

Japan Patent Office is not responsible for any damages caused by the use of this translation.

- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.**** shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

[Claim(s)]

[Claim 1] An object document acquisition means for a translation to acquire the becoming object document, and a non-translating phrase extract means to extract a non-translating phrase from the object document acquired with this object document acquisition means, It is the foreign language sentence reading comprehension support system which is equipped with a translation dictionary and a translation means to translate said object document using said translation dictionary, and is characterized by said translation means translating the text notation in said object document as a translation as a translation of the non-translating phrase extracted with said non-translating phrase extract means.

[Claim 2] An object document acquisition means for a translation to acquire the becoming object document, and a non-translating phrase extract means to extract a non-translating phrase from the object document acquired with said object document acquisition means, A notation conversion means to change into the notation by the word for a translation the non-translating phrase extracted with said non-translating phrase extract means, It has a translation dictionary and a translation means to translate said object document using said translation dictionary. Said translation means The notation changed by said notation conversion means as a translation of the non-translating phrase extracted with said non-translating phrase extract means, Or the foreign language sentence reading comprehension support system characterized by using both the notations of the notation changed by said notation conversion means, and the text notation in said object document, and translating said object document.

[Claim 3] Said non-translating phrase extract means is a foreign language sentence reading comprehension support system according to claim 1 or 2 characterized by what the noun phrase which does not exist independently in said translation dictionary is extracted for as a non-translating phrase.

[Claim 4] Said non-translating phrase extract means is a foreign language sentence reading comprehension support system according to claim 1 or 2 characterized by what a complex NP is extracted for as a non-translating phrase.

[Claim 5] It is a foreign language sentence reading comprehension support system given in any 1 claim of claim 1 to the claims 4 which are equipped with a significance judging means to judge the significance of said non-translating phrase, and are characterized by said translation means displaying said non-translating phrase in distinction from other translation parts according to the significance judged by this significance judging means.

[Claim 6] A retrieval means to search the similar document or the related description to the paragraph of the object document acquired by said object document acquisition means, a sentence, and at least one unit in a non-translating phrase, A foreign language sentence reading comprehension support system given in any 1 claim of claim 1 to the claims 5 characterized by providing the correlation means which associates the similar document searched with said retrieval means, or related description per relevance.

[Claim 7] The object document acquisition function for a translation which acquires the becoming object document, and the non-translating phrase extract function which extracts a non-translating phrase from the object document acquired by this object document acquisition function, It has a translation dictionary and the translation function to translate said object document using said translation dictionary. Said translation function The storage with which the foreign language sentence reading comprehension support program in which the computer read for realizing a computer is possible was memorized [translating the text notation in said object document as a translation as a translation of the non-translating phrase extracted by said non-translating phrase extract function, and].

[Claim 8] The object document acquisition function for a translation which acquires the becoming object document, and the non-translating phrase extract function which extracts a non-translating phrase from the object document acquired by said object document acquisition function, The notation conversion function to change into the notation by the word for a translation the non-translating phrase extracted by said non-translating phrase extract function, It has a translation dictionary and the translation function to translate said object document using said translation dictionary. Said translation function The notation changed by said notation conversion function as a translation of the non-translating phrase extracted by said non-translating phrase extract function, Or the storage with which the foreign language sentence

reading comprehension support program in which the computer read for realizing a computer is possible was memorized [using both the notations of the notation changed by said notation conversion function, and the text notation in said object document, and translating said object document, and].

[Claim 9] Said non-translating phrase extract function is the storage with which the foreign language sentence reading comprehension support program according to claim 7 or 8 characterized by what the noun phrase which does not exist independently in said translation dictionary is extracted for as a non-translating phrase was memorized.

[Claim 10] Said non-translating phrase extract function is the storage with which the foreign language sentence reading comprehension support program according to claim 7 or 8 characterized by what a complex NP is extracted for as a non-translating phrase was memorized.

[Claim 11] It is the storage with which it had the significance judging function to judge the significance of said non-translating phrase, and the foreign language sentence reading comprehension support program of a publication was memorized by any 1 claim of claim 7 to the claims 10 characterized by said translation function displaying said non-translating phrase in distinction from other translation parts according to the significance judged by this significance judging function.

[Claim 12] The retrieval function to search the similar document or the related description to the paragraph of the object document acquired by said object document acquisition function, a sentence, and at least one unit in a non-translating phrase, The storage with which the foreign language sentence reading comprehension support program of a publication was memorized by any 1 claim of claim 7 to the claims 11 characterized by providing the correlation function which associates the similar document searched with said retrieval function, or related description per relevance.

[Claim 13] The foreign language sentence reading-comprehension exchange approach characterized by to extract a non-translating phrase and to translate it from the object document used as the object for a translation using the notation of the non-translating phrase changed into the notation by the word for a translation as a translation of the extracted non-translating phrase, and one [at least] notation of the text tables in said object document in case said object document translates using a translation dictionary.

[Detailed Description of the Invention]

[0001]

[Field of the Invention] This invention relates to a foreign language sentence reading comprehension support system, the storage with which the foreign language sentence reading comprehension support program was memorized, and the foreign language sentence reading comprehension exchange approach, and relates to a translation of the document acquired in the detail.

[0002]

[Description of the Prior Art] The translation system which translates into the document of different language from it the document (foreign language sentence) drawn up in predetermined language like the system which translates into a Japanese document the document indicated in the foreign language is put in practical use. A translation system is widely used increasingly by current [which information can come to hand now to a large quantity freely by various communication links, such as the Internet and personal computer communications,]. In this translation system, the text (document for a translation) is translated into predetermined language according to a fixed algorithm using the dictionary for a translation. Since a translation is performed comparatively for a short time, it is effective, when communicating and you understand the homepage created in the foreign language.

[0003]

[Problem(s) to be Solved by the Invention] However, in the conventional translation system, if the phrase which does not exist in a dictionary was in the text, since the phrase would be translated literally, also when becoming the translation which is not well known as the whole sentence or becoming reading comprehension impossible occasionally, it was. Although this cannot recognize to accuracy phrases, such as a complex NP which should make semantics as one essentially, it is for making an abridged translation by separating these *****. For example, as shown in drawing 13 , in the conventional translation system, the part of "Action Manager" which does not exist in a translation dictionary will be translated as an "operation manager" by literal translation to the text of "Action Manager is now varing. So --." Consequently, now, "operation manager is diversified. In the decodelement therefore, --", I do not understand well in what thing.

[0004] Then, this invention was accomplished in order to solve said conventional technical problem, and it sets it as the 1st object to offer the foreign language sentence reading comprehension support system which can obtain a more natural translation result. Moreover, this invention sets it as the 2nd object to offer the storage with which the foreign language sentence reading comprehension support program which can obtain a more natural translation result was memorized. Moreover,

this invention sets it as the 3rd object to offer the foreign language sentence reading comprehension exchange approach which can obtain a more natural translation result. [0005]

[Means for Solving the Problem] In the foreign language sentence reading comprehension support system indicated to claim 1 An object document acquisition means 501 to acquire the object document which serves as an object for a translation as shown in drawing 5 , A non-translating phrase extract means 502 to extract a non-translating phrase from the object document acquired with this object document acquisition means 501, It has the translation dictionary 503 and a translation means 504 to translate said object document using said translation dictionary 503. Said translation means 504 Said 1st object is attained by translating the text notation in said object document as a translation as a translation of the non-translating phrase extracted with said non-translating phrase extract means 502. In the foreign language sentence reading comprehension support system indicated to claim 2 An object document acquisition means 501 to acquire the object document which serves as an object for a translation as shown in drawing 6 , A non-translating phrase extract means 502 to extract a non-translating phrase from the object document acquired with said object document acquisition means 501, A notation conversion means 505 to change into the notation by the word for a translation the non-translating phrase extracted with said non-translating phrase extract means 502, It has the translation dictionary 503 and a translation means 506 to translate said object document using said translation dictionary 503. Said translation means 506 The notation changed by said notation conversion means 505 as a translation of the non-translating phrase extracted with said non-translating phrase extract means 502; Or said 1st object is attained by using both the notations of the notation changed by said notation conversion means 505, and the text notation in said object document, and translating said object document. As shown in drawing 6 and drawing 7 , in invention indicated to claim 3, said non-translating phrase extract means 505 extracts the noun phrase which does not exist independently as a non-translating phrase in said translation dictionary 503 in the foreign language sentence reading comprehension support system indicated to claim 1 or claim 2. In invention indicated to claim 4, as shown in drawing 6 and drawing 7 , in the foreign language sentence reading comprehension support system indicated to claim 1 or claim 2, said non-translating phrase extract means 502 extracts a complex NP as a non-translating phrase. In invention indicated to claim 5, as one example is shown in drawing 7 , it has a significance judging means 507 to judge the significance of said non-translating phrase in a foreign language

sentence reading comprehension support system given in any 1 claim of claim 1 to the claims 4, and said translation means 504 displays said non-translating phrase in distinction from other translation parts according to the significance judged by this significance judging means 507. In invention indicated to claim 6, as one example is shown in drawing 7, it sets to a foreign language sentence reading comprehension support system given in any 1 claim of claim 1 to the claims 4. A retrieval means 508 to search the similar document or the related description to the paragraph of the object document acquired by said object document acquisition means 501, a sentence, and at least one unit in a non-translating phrase, The correlation means 509 which associates the similar document searched with said retrieval means 508 or related description per relevance is made to provide. The object document acquisition function 601 which acquires the object document which serves as an object for a translation as invention indicated to claim 7 shows to drawing 9, The non-translating phrase extract function 602 which extracts a non-translating phrase from the object document acquired by this object document acquisition function 601, It has the translation dictionary 603 and the translation function 604 to translate said object document using said translation dictionary 603. Said translation function 604 By using to translate the text notation in said object document as a translation as a translation of the non-translating phrase extracted by said non-translating phrase extract function 602 as the storage with which the foreign language sentence reading comprehension support program in which the computer read for realizing a computer is possible was memorized Said 2nd object is attained. The object document acquisition function 601 which acquires the object document which serves as an object for a translation as invention indicated to claim 8 shows to drawing 10, The non-translating phrase extract function 602 which extracts a non-translating phrase from the object document acquired by said object document acquisition function 601, The notation conversion function 605 to change into the notation by the word for a translation the non-translating phrase extracted by said non-translating phrase extract function 602, It has the translation dictionary 603 and the translation function 604 to translate said object document using said translation dictionary 603. Said translation function 604 The notation changed by said notation conversion function 605 as a translation of the non-translating phrase extracted by said non-translating phrase extract function 602, Or both the notations of the notation changed by said notation conversion function 605 and the text notation in said object document are used. Said 2nd object is attained by using to translate said object document as the storage with which the foreign language sentence reading comprehension support

program in which the computer read for realizing a computer is possible was memorized. As shown in drawing 9 and drawing 10, in invention indicated to claim 9, said non-translating phrase extract function 602 extracts the noun phrase which does not exist independently as a non-translating phrase in said translation dictionary in the storage with which the foreign language sentence reading comprehension support program indicated to claim 7 or claim 8 was memorized. In invention indicated to claim 10, as shown in drawing 9 and drawing 10, in the storage with which the foreign language sentence reading comprehension support program indicated to claim 7 or claim 8 was memorized, as for said non-translating phrase extract function 602, a complex NP is extracted as a non-translating phrase. In the storage with which the foreign language sentence reading comprehension support program of a publication was memorized in invention indicated to claim 11 by any 1 claim of claim 7 to the claims 10 as one example was shown in drawing 11. Having the significance judging function 607 to judge the significance of said non-translating phrase, said translation function 604 displays said non-translating phrase in distinction from other translation parts according to the significance judged by this significance judging function 607. It is ** about the similar document or the related description to the paragraph of the object document acquired from claim 7 by said object document acquisition function 601 in invention indicated to claim 12 in the storage with which the foreign language sentence reading comprehension support program of a publication was memorized by any 1 claim of the claims 11 as one example was shown in drawing 12, a sentence, and at least one unit in a non-translating phrase. The retrieval function 608 which carries out a funiculus, and the correlation function 609 which associates the similar document searched with said retrieval function 608 or related description per relevance are made to provide. In case said object document translates using a translation dictionary, a non-translating phrase carries out extract 701 and it takes translation 702 from the object document used as the object for a translation using the notation of the non-translating phrase changed into the notation by the word for a translation as a translation of the extracted non-translating phrase, and one [at least] notation of the text tables in said object document with the foreign language sentence reading-comprehension exchange approach which indicated to claim 13.

[0006]

[Embodiment of the Invention] Hereafter, the gestalt of the suitable operation in the foreign language sentence reading comprehension support system of this invention, the storage with which the foreign language sentence reading comprehension support program was memorized, and the foreign language sentence reading comprehension

exchange approach is explained with reference to drawing 9 from drawing 1.

(1) The foreign language sentence reading comprehension exchange by the outline book operation gestalt of an operation gestalt extracts predetermined non-translating phrases, such as a complex NP, from the documents for a translation (English etc.) by natural language processing first. And without translating the part of a non-translating phrase literally, it changes into a katakana notation and the whole object document is translated. The reading comprehension as the whole text which knows the content becomes easy by carrying out a katakana notation with an "action manager" by doing in this way, without translating "Action Manager" literally strangely. Moreover, even if it is the user who does not know the content, it can read and progress rather than an incomprehensible literal translation, without getting confused, since other parts can be read on the assumption that it is a special phrase.

[0007] (2) It not only constitutes from a computer system containing a personal computer, a word processor, etc., but the foreign language sentence reading comprehension support system of the detail book operation gestalt of the gestalt of operation can be constituted according to the computer system connected on the server of LAN (local area network), the host of a computer (personal computer) communication link, and the Internet. Moreover, it is also possible to carry out functional distribution to each device on a network, and to constitute a foreign language sentence reading comprehension support system from a whole network.

[0008] Drawing 1 is a block diagram showing the configuration of a foreign language sentence reading comprehension support system. The foreign language sentence reading comprehension support system is equipped with the control section 11 for controlling the whole system to be shown in drawing 1. The keyboard 12 as an input device, a mouse 13 and a display 14, an airline printer 15, storage 16, the storage driving gear 17, a communication controller 18, I/O I/F19, and a character reader 20 are connected to this control section 11 through the bus lines 21, such as a data bus. The control section 11 is equipped with CPU111, ROM112, and RAM113. ROM112 is the read only memory in which various programs and data for CPU111 to perform various control and an operation were stored beforehand.

[0009] RAM113 is random access memory used for CPU111 as a working memory. In this RAM113, as area for performing foreign language sentence reading comprehension exchange processing by this operation gestalt The object document storage area 1131 for a translation where the becoming object document is stored, The non-translating phrase extracted as a phrase which is not translated from an object document, and the non-translating phrase storage area 1132 which stores the

katakana notation, The various area of decodelement document storage area 1133 which stores the decodelement document translated into the non-translating phrase using the katakana notation, and others is secured.

[0010] A keyboard 12 constitutes a part of object document acquisition means in the case of inputting an object document and a divulgence word, and divulgence word input means within a self-system, and various keys, such as a function key for performing the kana key for inputting a kana alphabetic character, a ten key, and various functions and a cursor key, are arranged. A mouse 13 is a pointing device and is an input unit which specifies the function to correspond by left-clicking on the key displayed on the display 14, an icon, etc. As for an indicating equipment 14, CRT, a liquid crystal display, etc. are used. The input result by the keyboard 12 or the mouse 13 was displayed on this indicating equipment, or an object document, a decodelement document, etc. in foreign language sentence reading comprehension exchange processing were displayed on it, and it has come. An airline printer 15 is for printing the document displayed on the display 14, the document stored in the document storing section 164 of storage 16, and the bill to a user. As this airline printer, various airline printers, such as a laser beam printer, a dot impact printer, an ink jet printer, a page printer, a thermal printer, and a hot printing type printer, are used.

[0011] The store 16 consists of driving gears for writing various information, such as a program and data, to the storage which can be written, and its storage. Although a hard disk is mainly used, you may make it use the storage in which the R/W of the various storages used with the below-mentioned storage driving gear 17 is possible as a storage used for this store 16. The store 16 has the storing section (for example, storing section for backing up a program, data, etc. which are stored in this store 16) of the conversion-of-kana-into-kanji dictionary 161, the program storing section 162, the data storage section 163, the document database 164, the access hysteresis database 165, the translation dictionary 165 used for a translation, and others which are not illustrated etc. Various programs, such as a kana kanji conversion program which changes into a kanji mixture sentence the kana character string inputted using the conversion-of-kana-into-kanji dictionary 161 besides various programs, such as a foreign language sentence reading comprehension exchange processing program in this operation gestalt and a non-translating phrase extract program, are stored in the program storing section 162. The various data which systems, such as data about a user, need are stored in the data storage section 163. Many Japanese documents are mainly stored in the document database 164. A predetermined document etc. and a similar similar document are searched out of this Japanese document if needed.

Especially the format of each document stored in this document database 164 is not limited, and storing of the document of various formats, such as a document of text format, a document of a HTML (Hyper Text Markup Language) format, and a document of a JIS format, is possible for it.

[0012] The storage driving gear 17 is a driving gear for reading the data with which CPU111 contains a computer program and a document from an external storage. Various processing programs, such as foreign language sentence reading comprehension exchange processing performed by the foreign language sentence reading comprehension support system of this operation gestalt, and the dictionary used there, data, etc. are contained in the computer program memorized by the storage. A storage means the storage with which a computer program, data, etc. are memorized here. Specifically Magnetic storage media, such as a floppy disk, a hard disk, and a magnetic tape, Semi-conductor storages, such as a memory chip and an IC card, CD-ROM and MO, The storage using forms (and medium with the function equivalent to a form), such as storages with which information is read optically, such as PD (phase change rewriting mold optical disk), a punched card, and a paper tape, and the other storages computer programs etc. are remembered to be by various approaches are contained. As a storage used in the foreign language sentence reading comprehension support system of this operation gestalt, CD-ROM and a floppy disk are mainly used. The storage driving gear 17 can write in the data which a computer program is read from these various storages, and also are stored in RAM113 or a store 16 to the storage in which writing like a floppy disk is possible.

[0013] In the foreign language sentence reading comprehension support system of this operation gestalt, CPU111 of a control section 11 reads a computer program from the storage of the exterior set in the storage driving gear 17, and stores in each part of storage 16 (install). And when performing various processings, such as foreign language sentence reading comprehension exchange processing by this operation gestalt, from storage 16, an applicable program is read into RAM113 and performed. However, it is not from storage 16 and it is also possible to read a program into direct RAM113 from an external storage with the storage driving gear 17, and to perform. Moreover, ROM112 is made to memorize beforehand the foreign language sentence reading comprehension exchange processing program of this operation gestalt etc. depending on a foreign language sentence reading comprehension support system, and CPU111 may be made to perform this. Furthermore, the various programs and data of this operation gestalt, such as a foreign language sentence reading comprehension exchange processing program, are downloaded from other storages

through a communication controller 18, and it may be made to perform them.

[0014] A communication controller 18 is a control unit for carrying out network connection of between a foreign language sentence reading comprehension support system and various electronic equipment, such as other personal computers and a word processor. It was inputted and also it is possible the document of the language as the object document which these various electronic equipment has with the same communication controller 18, and to access the database of the document of language, and the document of the same language or a divulgence word as an object for retrieval. Various data, such as bit map data besides the document of various formats, such as text format and a HTML format, are also contained in the target document. I/O I/F19 is an interface for connecting various devices, such as a loudspeaker which outputs voice, music, etc. A character reader 20 is equipment which recognizes the alphabetic character indicated by the form etc. in various formats, such as text format and HTML, and consists of an image scanner, a character reader, etc.

[0015] It is created by the alter operation of a keyboard 12 with this operation gestalt. Others [document / (it stores in the predetermined storage area of RAM113)], The document drawn up in the exterior read from the storage with the storage driving gear 17, What (document acquisition means) is acquired as an object document set as the object of a translation of various documents, such as a document beforehand stored in the document database, a document downloaded from the communication controller 18, and a document which carried out character recognition with the character reader 20, is possible.

[0016] Actuation of the foreign language sentence reading comprehension exchange processing by the foreign language sentence reading comprehension support system of this operation gestalt constituted as mentioned above is explained from drawing 2 using drawing 4. With this operation gestalt, the document (text) which wishes to translate is indicated in English, and the case where this is translated into Japanese is explained to an example.

[0017] Drawing 2 is a flow chart showing the Maine actuation of foreign language sentence reading comprehension exchange processing. CPU111 acquires the object document (text) which wishes foreign language sentence reading comprehension exchange (translation) first (step 11).

[0018] And the language (for example, English) of the text specified with the keyboard 12 or the input unit of mouse 13 grade, an interpretive language (for example, Japanese), and a display format are acquired, and it stores in the predetermined storage area of RAM113 (step 12). Here, when assignment of text language is not

done by the user, CPU111 analyzes the acquired object document and specifies language. Analysis is performed with checking existence of the notation alphabetic character of each language proper. Moreover, since the own alternative of an interpretive language is an indispensable item in principle, when selection is not made, selection is urged to it by the screen display. However, when a foreign language sentence reading comprehension support system is possible only for one kind of translation (for example, only English to Japanese), acquisition and storing of text language and an interpretive language are unnecessary. When acquiring text language, an interpretive language, and a display format, CPU111 displays the combination of the language which can be translated, and a display format on a screen, and it is made to have a user choose it.

[0019] Next, CPU111 stores in the non-translating phrase storage area 1132 the non-translating phrase (notation of the text) which started the PU non-translating phrase extract program corresponding to text language, carried out the resource language-processing (NLP) extract of the non-translating phrase, and was extracted from the object document stored in the object document storage area 1132 of RAM113 (step 13). Here, as an example of the phrase extracted as a non-translating phrase, a complex NP is extracted and there are a noun + pronoun, a noun + proper noun, a proper noun + article + common noun (example; Billy the kid), an adjective + common noun + proper noun (example; Mad scientist Doctor Emett Brown), etc. as the example. Furthermore, if it illustrates, when text language will be English, complex NPs, such as "An incremental utterance taskoriented dialogue production system" and "An Internet based wide area computer assisted foreign language learning prototype demonstration system", are extracted as a non-translating phrase.

[0020] Next, CPU111 checks whether it is the display format in which the display format specified at step 12 includes a katakana display (step 14). The katakana notation which changed into the katakana notation the non-translating phrase extracted at (;Y) and step 13 when it was a display format including a katakana display (notation conversion means), was made to correspond to a non-translating phrase and was changed is stored in the non-translating phrase storage area 1132 (step 15). Here, conversion for the katakana notation of a non-translating phrase is changed by Roman alphabet reading or the phonetic symbol of each word, when the text is English.

[0021] And CPU111 translates an object document using the translation dictionary 165, and the non-translating phrase and/or katakana notation (it chooses according to the selected display format) which were stored in the non-translating phrase storage area 1132, and stores a translation result in the decodement document

storage area 1133 (step 16). After a translation is completed, CPU111 is displayed by the display format which had the decodement document stored in RAM113 specified (step 17), and ends processing.

[0022] Drawing 3 expresses the display screen of a translation result when a "katakana notation" is chosen as a display format. In this drawing 3, although the text (object document) and a Japanese translation result are displayed as contrasted with width, it is also possible to display only either, or to contrast up and down and to display. Since a non-translating phrase "Action Manager" was not strangely translated literally into the incomprehensible word of an "operation manager", but it bundled up with the "action manager" and it has written in katakana as shown in this drawing, the user who knows the language concerned can understand promptly. Moreover, even if it is the user who does not know the language concerned, it can read and progress, without getting confused, since it cannot say that the whole text is not clear anymore by incomprehensible literal translation, and other parts can be read on the assumption that it is a special phrase.

[0023] Drawing 4 expresses the display screen when other display formats are chosen, (A) expresses the display screen of a "text notation" and (B) expresses the display screen of "a katakana notation (text notation)." As shown in this drawing, by carrying out a text notation, for a user with the linguistic force of that text, he can understand the content promptly from a text notation, and the word which must be investigated in a dictionary etc. from a text notation with the user who is not so can be clarified. In addition, although not illustrated, when "a text notation (katakana notation)" is chosen as a display format, it is displayed so much on a non-translating phrase "Action Manager" as "Action Manager (action manager)."

[0024] As mentioned above, although the processing which translates katakana notation / text notation mixture to the object document by this operation gestalt was explained, it may be made to perform processing of further the following. That is, the significance of the non-translating phrase extracted at step 13 is judged using an object document, according to the significance, others carry out partial special [of the text notation or katakana notation part in the decodement of a translation result], and they are displayed. Using alphabetic character decoration information, such as a size of a color, concentration, a character size, and an alphabetic character, etc. as a distinction display of a text notation or a katakana notation, it emphasizes or may be made to carry out reverse emphasis. By the case where significance is high, red etc. emphasizes emphasizing by conspicuous color and high concentration, the enlarged character (magnitude is changed according to significance), *****, and any one or

more **. By the case where significance is low, a light blue, gray, etc. carry out reverse emphasis of carrying out reverse emphasis so that it may not be conspicuous one or more [of color and low concentration, reduced-character, and thin alphabetic character **]. [which are not conspicuous]

[0025] As the judgment approach of the significance of a non-translating phrase, the well-known important word judging approach by which various proposals are made is used. For example, PU111 extracts an independent word from an object document by performing morphological analysis, and is taken as candidate language. And from the frequency of occurrence in the inside of the object document of the extracted candidate language, and a performance index, the significance of each candidate language is determined and let the candidate language which has the significance beyond a fixed threshold be an important word. Here, as a performance index, when the predetermined phrase is specified beforehand, weighting to the phrase etc. is used, for example. And the important word candidate who appears in a non-translating phrase is checked, and let total value of significance be the significance of the non-translating phrase.

[0026] Moreover, the important word in a document is extracted from information, such as a type of letters and a verb, and you may make it judge it as the judgment approach of the significance of a non-translating phrase, as proposed by "an extract and processing" (the collection of the 38th time national conference drafts of Information Processing Society of Japan, the 222nd page; 1989) of the semantic information from whole sentence information.

[0027] Moreover, in addition to the processing which translates katakana notation / text notation mixture to the object document by this operation gestalt, it may be made to perform processing of further the following. that is, the similar document or the related description boiled and depended is searched for every ** [at least] of the paragraph of the object document acquired at step 11, a sentence, and the non-translating phrases, and a hyperlink is attached. And when a user scrolls in the part concerned or holds up a mouse cursor, a part of Applicable-documents name list and content of a document (a head and epitome) are displayed in the shape of balloon (blowdown), and an understanding is supported. Under the present circumstances, it may be made to indicate the Applicable documents of a native language by precedence.

[0028] Here, retrieval of a similar document uses the document vector (vector which uses as an element significance of the keyword contained in a document) which can use various well-known search methods, for example, expresses the description of a

document, and asks for the similarity of both documents as follows by the cosine depending on the include angle between the document vectors over two documents. That is, when the include angle between both the documents vector b_n and b_{n+1} is set to q , the inner product of both the documents vector is made into $b_n \cdot b_{n+1}$ and magnitude of both the documents vector is made into $|b_n|$ and $|b_{n+1}|$, respectively, similarity S_{n+1} of both the documents vector can be found with the following formula 1.

[0029]

[Equation 1]

Similarity $S_{n+1} = \cos(q)$

$= b_n \cdot b_{n+1} / (|b_n| \cdot |b_{n+1}|)$

[0030] The value of this similarity S takes the value to $-1 \leq S \leq 1$, and two document vectors can consider mutually that near and two documents are alike to parallel, so that it is close to 1. Therefore, the document which has the similarity beyond a fixed threshold is judged to be a similar document, and it considers as a retrieval result.

[0031] As mentioned above, although the configuration of this operation gestalt and processing of foreign language sentence reading comprehension exchange were explained, it is possible to carry out deformation various by within the limits of invention which is not limited to each of these gestalten and indicated by the claim by this invention. For example, although it explained to English as language (text language) of an object document and the operation gestalt explained Japanese to the example as an interpretive language, it is possible for it not to be limited to such combination in this invention, and to consider as the combination between all language. In that case, what is necessary is just to change an uninfluential part into the configuration of this invention of using the morphological analysis algorithm for language with which the target document was drawn up.

[0032] It is possible to constitute an operation gestalt with each means (- means) as a superordinate concept which contains them to each equipment, each part, each actuation, each processing, etc. explained in the above operation gestalt. For example, you may make it constitute a "similarity calculation means" to the publication "2 Ask for the similarity of both documents as follows by the cosine depending on the include angle between the document vectors over a document." You may make it similarly constitute an operation gestalt from superordinate concepts, such as "- (actuation) means", to various actuation.

[0033]

[Effect of the Invention] In case said object document is translated using a translation

dictionary according to this invention, from the object document used as the object for a translation, extract a non-translating phrase and as a translation of the extracted non-translating phrase Since it was made to translate using the notation of the non-translating phrase changed into the notation by the word for a translation, and one [at least] notation of the text tables in said object document, a more natural translation result is offered and foreign language sentence reading comprehension is supported.

[Brief Description of the Drawings]

[Drawing 1] It is a block diagram showing the configuration of the foreign language sentence reading comprehension support system in 1 operation gestalt of this invention.

[Drawing 2] It is a flow chart showing the Maine actuation of the foreign language sentence reading comprehension exchange processing in the same as the above and an operation gestalt.

[Drawing 3] It is an explanatory view showing the display screen of a translation result when a "katakana notation" is chosen as a display format in the same as the above and an operation gestalt.

[Drawing 4] It is a thing showing the display screen when other display formats in the same as the above and an operation gestalt are chosen, and is the explanatory view in which (A) expressed the display screen of a "text notation" and (B) expressed the display screen of "a katakana notation (text notation)."

[Drawing 5] It is drawing corresponding to a claim of invention indicated to claim 1.

[Drawing 6] It is drawing corresponding to a claim of invention indicated to claim 2.

[Drawing 7] It is one example of drawing corresponding to a claim of invention indicated to claim 5.

[Drawing 8] It is one example of drawing corresponding to a claim of invention indicated to claim 6.

[Drawing 9] It is drawing corresponding to a claim of invention indicated to claim 7.

[Drawing 10] It is drawing corresponding to a claim of invention indicated to claim 8.

[Drawing 11] It is one example of drawing corresponding to a claim of invention indicated to claim 11.

[Drawing 12] It is one example of drawing corresponding to a claim of invention indicated to claim 12.

[Drawing 13] It is drawing corresponding to a claim of invention indicated to claim 13.

[Drawing 14] It is an explanatory view showing the conventional translation result.

[Description of Notations]

- 11 Control Section
- 112 ROM
- 113 RAM
- 1131 Object Document Storage Area
- 1132 Non-Translating Phrase Storage Area
- 1133 Decodement Document Storage Area
- 12 Keyboard
- 13 Mouse
- 14 Display
- 15 Airline Printer
- 16 Storage
 - 161 Conversion-of-Kana-into-Kanji Dictionary
 - 162 Program Storing Section
 - 163 Data Storage Section
 - 164 Document Database
 - 165 Translation Dictionary
- 17 Storage Driving Gear
- 18 Communication Controller
- 19 I/O I/F
- 20 Character Reader

[Translation done.]

* NOTICES *

Japan Patent Office is not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. *** shows the word which can not be translated.
3. In the drawings, any words are not translated.

DRAWINGS

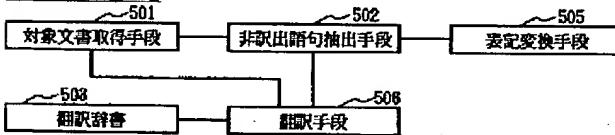
[Drawing 3]

原文	日本語
Action Manager is now varying. So	アクション マネージャーは、今や多様化してい る。 従って、.....
⋮	⋮

[Drawing 5]



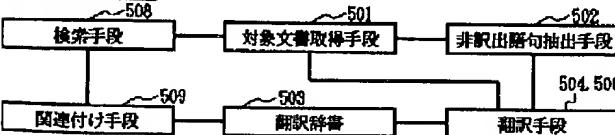
[Drawing 6]



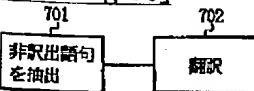
[Drawing 7]



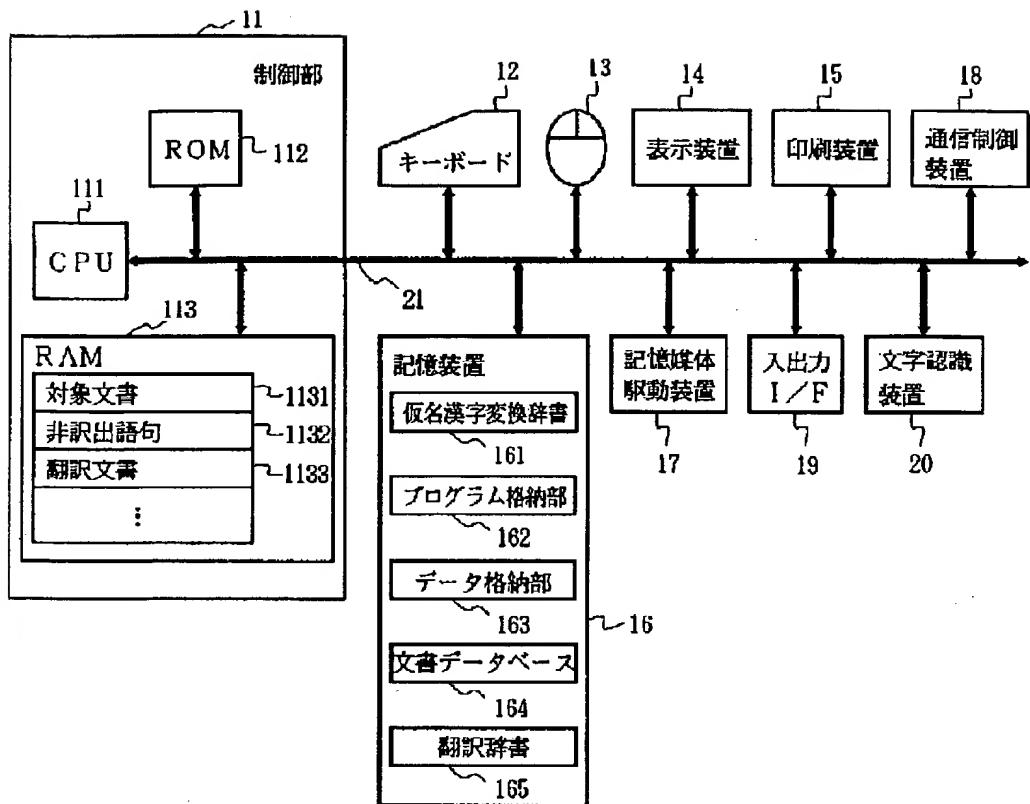
[Drawing 8]



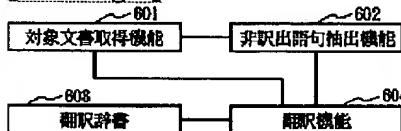
[Drawing 13]



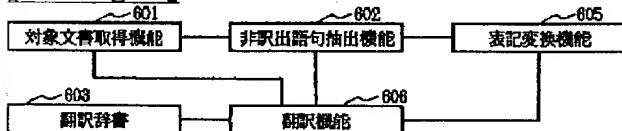
[Drawing 1]



[Drawing 9]



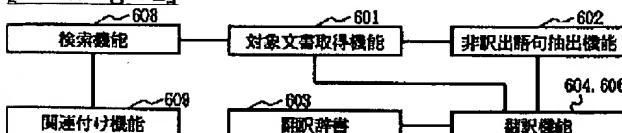
[Drawing 10]



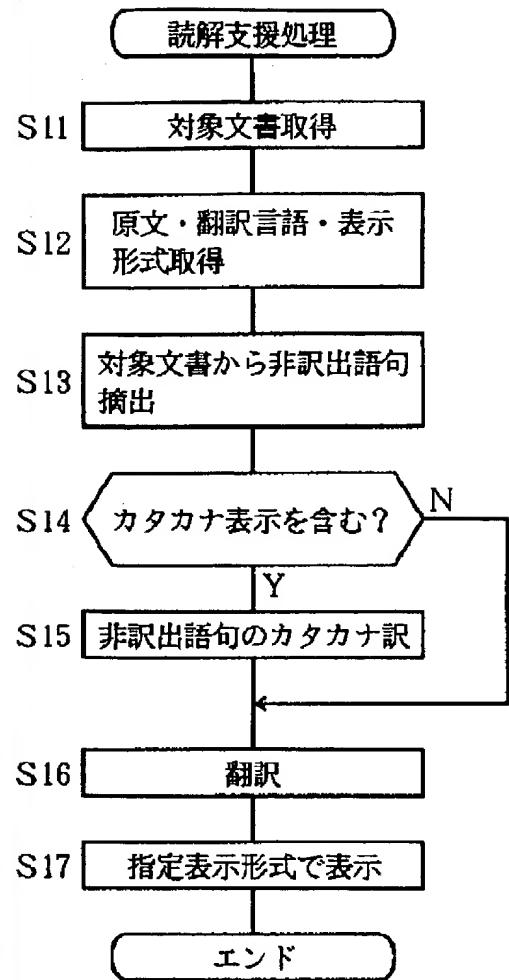
[Drawing 11]



[Drawing 12]



[Drawing 2]



[Drawing 4]

(A)

原文	日本語
Action Manager is now varing. So ⋮	Action Managerは、今や 多様化している。 従って、..... ⋮

(B)

原文	日本語
Action Manager is now varing. So ⋮	アクション マネージャー(Action Manager)は、 今や多様化している。 従って ⋮

[Drawing 14]

原文	日本語
Action Manager is now varing.	演技課長は、今や多様化 している。
So	従って、.....
...	...

[Translation done.]